

# Assessment practices of an MBA degree program

Gryphon Sou

Asia International Open University (Macau)

## Abstract

**Background:** From 2007 to 2008, serial researches were conducted on the Student Learning Experiences, Student Learning Outcomes and Assessment Practices of an MBA degree program offered by the Asia International Open University (Macau) in collaboration with the higher institutions of the Mainland.

**Aims:** This paper reports the findings of a research on the assessment practices of the subject program.

**Sample:** This research involved six internal examiners, nine external examiners and two registrars of the university offering the subject program. It also involved 331 MBA candidates from 20 higher institutions of mainland China who completed the subject program.

**Method:** Qualitative data from Technical and Non-Technical Literature was processed by Grounded Theory while quantitative data was analyzed with the aid of One Way ANOVA, Dunnett's tD Test, S-N-K Test and Tukey Test.

**Results:** Qualitative and quantitative data analysis disclosed that variances of a common assessment task subsisted in the subject program. The less favorable Student Learning Experiences of the sampled candidates were partially attributed to the marking variances of the thesis supervisors and the examiners.

**Conclusion:** The thesis supervisors regarded the assessment task of marking the thesis as a Norm-Referenced Assessment while the examiners regarded it as a Criteria-Referenced or an Objective-Referenced Assessment. This area of Assessment Practices is worth further study.

**Keywords:** Assessment Practices; Criteria-Referenced Assessment; Norm-Referenced Assessment; Objective-Referenced Assessment; Student Learning Experiences; Student Learning Outcomes.

# 工商管理碩士課程的考核手段

仇志成

亞洲（澳門）國際公開大學

## 摘要

**背景：**自2007至2008年筆者在亞洲（澳門）國際公開大學進行了一系列有關工商管理碩士課程的「學習體驗」、「學習成果」及「考核手段」的研究。

**目的：**本文旨在報告就亞洲（澳門）國際公開大學與國內高等院校合辦有關課程的考核手段所作研究的初步結果。

**調查對象：**是次研究涉及提供有關課程的大學中6名校內主考、9名校外主考及2名教務長，同時涉及國內20所高等院校及331名完成有關課程的考生。

**調查方法：**以紮根理論法對「技術文獻」及「非技術文獻」作定性數據分析；而定量數據分析則通過單向「變異數分析法」、「杜奈特檢定法」、「紐曼－柯爾差距檢定法」及「杜凱氏事後比較法」加以處理。

**調查結果：**定性及定量數據分析表明有關課程的「考核手段」邏輯上存在差異，而受查考生較為不悅的「學習體驗」部份亦歸咎於論文導師及主考在評分上的差異。

**總結：**得出的初步結論是論文導師以「定額基準評核」作為有關課程的「考核手段」，而主考則以「基準為本評核」或「目標為本評核」作為是項課程的考核手段。總括而言，有關課程的「考核手段」值得進一步研究。

**關鍵詞：**評核手段、基準為本評核、定額基準評核、目標為本評核、學習體驗、學習成果。

## 1. Synopsis

This paper reports the findings of a research on the assessment practices of a Master of Business Administration (MBA) degree program offered by Asia International Open University (Macau) (AIOU) in collaboration with the higher institutions of the Mainland. The subject program is mainly taught by the Mainland scholars while the candidates are examined by the AIOU and its external faculty staff. For graduation, the candidates are required to complete an action research project under the supervision of a Mainland scholar. The research project leading to a master thesis will be respectively marked by an internal and an external examiner. Then, the candidates have to attend a viva voce before

the internal and the external examiners for the award.

From 30.01.07 to 24.04.08 (sampling period), 23 Focus Group Interviews (FGI) were held in Macau Special Administrative Region (Exhibit 1). Voluntary participation of the FGI eventually exhausted all the available six internal examiners, nine external examiners and two registrars of the AIOU. One of the registrars was a doctoral candidate while the other participants of the FGI were earned philosophy doctorates or professional doctorates<sup>1</sup>. The external examiners included academics from the other universities or management practitioners in the commerce, industry or public services.

Exhibit 1: 331 MBA Candidates assessed and 23 Focus Group Interviews from 30.01.07 to 24.04.08

S/N	Dates	MBA Candidates assessed	Higher Education Institutions <sup>2</sup> of PRC	Participants of Focus Group Meetings		
				Internal Examiners	External Examiners	Registrars
1	30.01.07	13	1	2	5	2
2	31.01.07	6	1	2	5	2
3	14.03.07	16	2	2	5	2
4	18.04.07	8 + 11	1 + 3	2	4	2
5	19.04.07	7 + 9	4 + 5	5	5	2
6	22.05.07	8 + 5	6 + 7	3	3	2
7	21.06.07	15	8	3	5	1
8	22.06.07	7 + 8	5 + 9	3	5	1
9	18.07.07	17	10	4	6	2
10	19.07.07	16	11	4	6	2
11	02.08.07	7 + 5	12 + 7	3	5	2
12	30.10.07	13 + 2	13 + 14	2	4	2
13	31.10.07	8 + 10	7 + 15	2	3	1
14	20.11.07	8	16	2	5	2
15	05.12.07	14	10	4	3	2
16	11.12.07	15 + 1	17 + 18	3	4	1
17	13.12.07	8 + 7	19 + 20	4	4	2
18	15.01.08	15	6	2	4	2
19	16.01.08	14 + 1	17 + 13	3	4	2
20	09.04.08	15	8	1	4	2
21	10.04.08	8 + 6	13 + 8	3	4	2
22	23.04.08	14	6	2	3	2
23	24.04.08	8 + 6	5 + 15	3	2	1
23 days		331 MBA candidates	20 higher institutions	64 times	98 times	41 times

(Source: The subject research)

1. Doctor of Education, Doctor of Letters, Doctor of Literature, Doctor of Engineering, Doctor of Management or Doctor of Business Administration.

The registrars organized the MBA candidates to appear before the academic panels for viva voce. The internal and external examiners formed various academic panels which assessed all the MBA candidates of AIOU in the sampling period. Assessment forms of 331 MBA candidates from 20 higher institutions of the Mainland were selected for further study (Exhibit 1). The higher education institutions involved were traditional universities, private colleges and national training establishments for the comrades in mainland China.

## 2. Research Methodology

Under the Grounded Theory Approach (GTA), the data from “Technical or Non-Technical Literature” went through open or axial or selective coding and Constant Comparison Processes (CCP). The “Technical and Non-Technical Literature” involved in the data analysis are tabulated as per Exhibit 2:

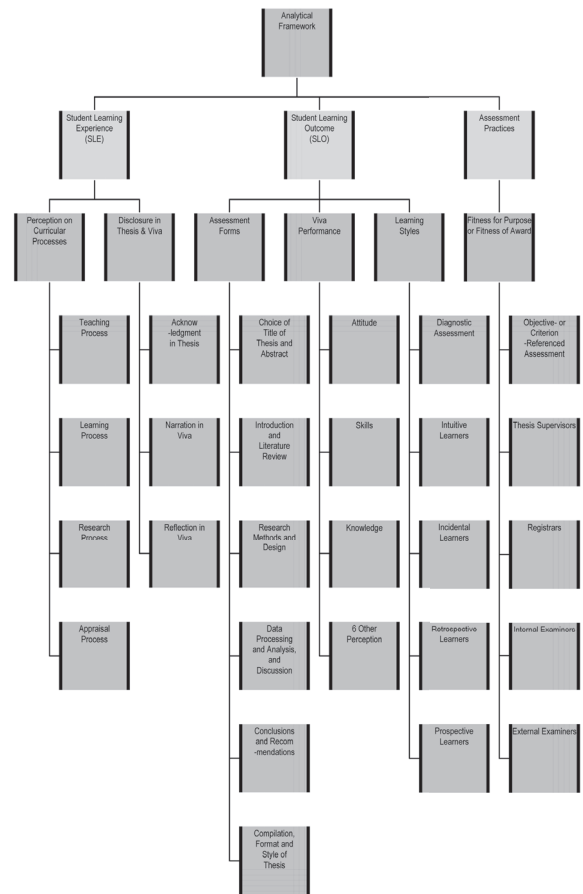
Exhibit 2: “Technical Literature” and “Non-Technical Literature” used in Data Analysis

<b>Technical Literature</b> (Secondary Data)	<ul style="list-style-type: none"> <li>✧ Archives of AIOU</li> <li>✧ Reports on previous researches in China</li> <li>✧ Publications of referential value to the subject research</li> </ul>
<b>Non-Technical Literature</b> (Primary Data)	<ul style="list-style-type: none"> <li>✧ 331 Assessment Forms of MBA Candidates</li> <li>✧ 331 MBA Theses of the MBA Candidates from 20 Higher Education Institutions.</li> <li>✧ Field-Notes of Focus Group Interviews</li> <li>✧ Memos with Theoretical Notes, Operational Notes and Code Notes under the Grounded Theory Approach</li> </ul>

(Source: The subject research)

Eventually, there were 3 core-categories, 6 categories and 27 sub-categories generated in the course of data analysis (Exhibit 3). In this paper, the core-category, categories and sub-categories pertaining to “Assessment Practices” will be analyzed independently or communally.

Exhibit 3: Diagram illustrating the Analytical Framework



(Source: The subject research)

<sup>2</sup>1. Ex-Tianjin Zhong Xin International Institute of Further Studies (前天津中新國際進修學院); 2. Jiang Han University (江漢大學); 3. Changsha Han Shuo Academy of Management (長沙漢碩管理專修學院); 4. Guangdong Academy of Technology for Comrades (廣東省科技幹部學院); 5. Shenzhen Hua Lian College of Commerce and Industry (深圳市華聯工商進修學院); 6. Guangxi Academy of Economics and Management for Comrades (廣西經濟管理幹部學院); 7. Guangdong Academy of Economics and Management for Comrades (廣東經濟管理幹部學院); 8. Liaoning Academy of Economics and Management for Comrades (遼寧經濟管理幹部學院); 9. Guangdong Academy of Technology for Comrades (廣東省科技幹部學院); 10. Northwest University (西北大學); 11. Beijing Academy of Finance and Management for Comrades (北京市財貿管理幹部學院); 12. Tianjin Academy of Finance and Management for Comrades (天津市財貿管理幹部學院); 13. Chengzhou Center of Adult Education (鄭州市成人教育中心); 14. Shandong Tai Shan College of Management (山東泰山管理學院); 15. China Advanced Studies and Research University (中華研修大學); 16. National Finance Commission Training Center for Comrades (國資委幹部教育培訓中心); 17. Tianjin Faculty of Business (天津商學院); 18. Anhui University (安徽大學); 19. National Economics and Trade Commission Shandong Office of Occupational Education (國資委幹部教育培訓中心); 20. Tianjin Academy of Finance and Trade for Comrades (天津財貿管理幹部學院).

### 3. Data Analysis under Grounded Theory Approach

To recap, the core category pertaining to Assessment Practices (Exhibit 3) has 1 category called “Fitness for Purpose or Fitness of Award” and 5 sub-categories, namely (1) Objective/Criterion-Referenced Assessment, (2) Thesis Supervisors, (3) Registrars, (4) Internal Examiners, and (5) External Examiners. In the following sections, these category and sub-categories will be analyzed through study of “Technical Literature” and coding of “Non-Technical Literature”.

In implementing effective assessment practices, there will be a grave concern over the issues of fairness, reliability and validity (Brown et al., 1997). In the review of “Technical Literature”, these issues can be construed as follows (Exhibit 4):

Exhibit 4: Three Issues in Assessment Practices

Fairness	Equality of opportunity and treatment.
Reliability	Consistency of approach.
Validity	Appropriateness of methods of truth-seeking.

(Source: The subject research)

“Fairness”, “Reliability” and “Validity” of assessment practices are deep problems. Atkins et al. (1993) criticize the procedures of many universities for student assessment. Barnett (1994) further criticizes the conflicts of purpose, variability across and within subject assessments, the sample of activities assessed, the methods of assessment and the grading system of worth in the English test system. With such criticism, is it possible to implement assessment practices in an effective way? Furthermore, is it possible to seek the truth of an educational system as questioned by Rowntree (1987, p1)?

*If we wish to discover the truth about an educational system, we must look into its assessment procedures. What student qualities and achievements are actively valued and rewarded by the system? How are its purposes and intentions realized?*

In this research, review of “Non-Technical Literature” revealed that the candidates had relatively less favorable SLE in the viva. In the FGI, the registrars reported that the candidates had actually commented on the “fairness” and the “reliability” of the viva. Occasionally, some candidates came up to the registrar and queried about the unfavorable results in the viva or solicit her assistance in re-submission of the theses for grading. They perceived variability across the academic panels and the thesis supervisors, the methods of assessment and the grading system of worth. When the candidates from the same institution were assigned to different academic panels for the viva, they might get different assessments from the examiners.

This phenomenon was thoroughly investigated in this research. The investigation started with an exploration of “Technical Literature” in terms of “fairness”, “reliability” and “validity” of effective assessment practices. To this end, the nature of “reliability” and “validity” and their implications for assessment practices will be examined. In this paper, statistical techniques will be used in a simple way as they are based essentially on measures of agreements and differences and range in complexity from correlations and analyses of variance to factor analyses, multi-variate analysis and beyond (Ebel & Frisbie, 1986; Grounlund, 1988). Instead, focus of study will be laid on the underlying concepts that are crucial to “fair” and effective assessment practices.

The standard approaches to reliability and validity are derived from psychometrics. They are based on the notion of an ideal which can be achieved if only one can reduce the errors. In management education, there is a range of values involved at the higher levels of abilities, skills and knowledge of which their integration could hardly assume that there is only one ideal. Hence, non-statistical approaches such as the use of judgments in the sampling of research project, definition of research problems, identification of the burning issues, application of research methods and proposition of remedial action are required in the assessment practices of management education. Nonetheless, non-statistical judgmental approaches of student assessment are also based on the underlying concepts of reliability and validity – notions of precision and accuracy.

Brown et al. (1997, p. 234) tell an analogical story about “watch and time”. The mechanism of a watch may be precise (reliable), it may measure the minutes and hours consistently, but the time show may be wrong. The time shown on a particular occasion may be correct (valid) but the watch may have stopped or its variable rate of loss or gain rather than its consistency may have provided the result. Furthermore, the watch-keeper may also be a variable when he or she fails to read and interpret the watch dial correctly.

The above analogy can be applied to interpreting the results of the assessment practices in the subject program. Even if guidelines and assessment forms are provided, ultimately the assessment instrument is the examiner in conjunction with the particular guidelines, assessment forms and procedures. In analyzing data about the assessment practices, the

focus can be placed upon the fairness, reliability and validity of the assessment task and its actual nature. Considering “fitness for purpose” and “fitness of award”, the thesis supervisors, the registrars, the internal and external examiners should be able to tell whether the assessment practices of the subject program are effective or not.

### 3.1 Assessment Practices

With reference to “Technical Literature”, taking a sample of what the candidates do, making inferences and estimating the worth of their actions (Brown et al., 1997) may be regarded as effective assessment practices. Obviously, assessment practices of the subject program equates to:

Sampling of Candidates + Making Inferences +  
Estimating Worth of Actions

First, sampling is undertaken by the candidates themselves, their thesis supervisors or even their employers. It involves the learning tasks: selecting a research topic, identifying the burning issues, conducting an action research, reporting the findings, proposing the solutions, and writing a thesis. After sampling, inferences could be made from the thesis and through the viva about the SLO such as Attitude, Skills, Knowledge, Achievements, Potential, Intelligence, Aptitudes, Motivations, Personality (Brown, et al., 1997). With the inferences, the examiners estimate the worth of the candidates’ actions. The estimation is in the form of grades, marks, recommendations.

These 3 aspects of assessment task as learning task have respective shortcomings which lead to problematic assessment practices. **Sampling** may

not be representative of the candidate's capabilities or may not match the learning objectives of the subject program. It may be drawn on too narrow a domain, such as one of the core subjects. Besides, it may be over-weighted towards particular skills or methods instead of applying integrated skills and knowledge gained from the subject program.

The **Inferences** drawn about the candidate's research may vary widely from examiner to examiner. The variations may be more significant when explicit criteria or marking schemes are not used. The **Estimation of Worth** in terms of the marks or grades may also vary. In the subject program, the variation in grading and marking would even lead to unjustifiable decisions by the academic panels.

Student assessment can be based upon the procedure of "Sampling", "Making References" and "Estimating Worth". On this basis, assessment practices would carry some common weaknesses (Brown, et al., 1997, pp 251-52):

- ✧ The sample does not match the stated outcomes.
- ✧ The sample is drawn from too narrow a domain.
- ✧ The sample is too large or too small.
- ✧ Absence of well-defined criteria.
- ✧ Unduly specific criteria.
- ✧ Variations in the inferences drawn by different assessors of the sample.
- ✧ Variations in estimates of worth.

Data analysis of this research proves and disproves some of the above weaknesses. During the FGI, the registrars, internal and external examiners

concurred that an action research was an appropriate sample with well-defined and specific criteria. The candidates were required to draw a live problem from the domain of business administration and applied their knowledge and skills gained from the subject program to solve the problem. The candidates proposed solutions in their theses from which the examiners drew inferences how the candidates applied their skills and knowledge gained from the subject program.

In estimating the worth of the proposed solutions, internal and external examiners were supposed to adopt the well-defined and specific criteria. The internal and external examiners reflected that they were committed to a fair, reliable and valid assessment. However, some of them noticed that there were variations in the inferences drawn by different examiners as well as different panels. Occasionally, the internal examiners and the external examiners could not compromise in the final marks of the candidates. Some opined that the existing assessment practices were most effective while some said that there were rooms for improvement.

### **3.1.1 Benefits and Shortcomings of Assessment Practices**

Effective assessment practices can be beneficial to the educational quality assurance. However, assessment practices established on sampling, inferences and estimation appear to be problematic. Brown et al. (1997) list some shortcomings of assessment practices which could be compared with the findings of this research (Exhibit 5):



Exhibit 5: Shortcomings of Assessment Practices vis-à-vis Research Findings

Item	Shortcomings of Assessment Practices	Findings of this Research
1	Overload of the candidates and the examiners.	Registrars reported that the candidates had to attend a pre-viva in the Mainland and a viva in Macau. Internal and External Examiners said that the assessments of theses were tedious.
2	Insufficient time for the candidates to complete the action research in the time available.	Some candidates reflected that they wished to do a big research project but time did not allow it.
3	Insufficient time for the examiners to mark the theses before the viva.	Some internal and external examiners reported that marking of a single thesis might take an hour before the viva. Some thesis might take longer time to digest.
4	Inadequate or superficial feedback provided to the candidates.	Some internal and external examiners responded that they wished to provide more feedback to the candidates during the half-an-hour viva. They assessed from the candidates' performance during the viva that the feedback from the thesis supervisors to a candidate during the research appeared to be inadequate and superficial.
5	Wide variations in assessment demands of different panels.	Internal and external examiners noted that different combination of the panel members varied in their assessment demands.
6	Wide variations in marking across panels.	Internal and external examiners noted that there were wide variations across different panels.
7	Wide variations in marking within a panel.	Some internal and external examiners noted that there were variations about 10 scores in their markings before the viva. However, they could promise with each other during the viva.
8	Wide variations in marking by supervisors.	Internal and external examiners commented that there were variations in marking by supervisors. Generally, their markings were higher than that by the panels.
9	Fuzzy or non-existent criteria.	Internal and external examiners believed that criteria existed but they were subject to their own interpretation.
10	Undue precision and specificity of marking schemes or criteria.	Internal and external examiners commented that precise and specific marking schemes or criteria might not be viable for the action research in the management education.
11	Candidates do not know what is expected of them.	Internal and external examiners commented that some candidates did not know the exact requirements of the action research as well as the thesis.
12	Candidates do not know what counts as a good or bad research or thesis.	Internal and external examiners commented that some candidates had not realized the strengths and weaknesses until they were enlightened during the viva.

(Source: The subject research)

With an array of shortcomings, it is questionable whether the assessment practices could assure educational quality. What supports the use of assessment practices to assure quality education? It is believed that assessment supports learning and assessment task should be regarded as a learning task. This belief is supported by a number of scholars over the years. Hattie and Watkins (1985) comment that using projects and open-ended assessments tend to promote independence and deeper strategies of learning.

Though using problem-based approaches and appropriate research projects tends to promote

deeper styles of learning, deeper approaches to study and independent learning tend to decline in higher education (Bain & Thomas, 1984; Biggs, 1987; Clarke & Newble, 1987; Eysenck, et al., 1987; Harper & Kember, 1989; Blake & Vernon, 1994). The declining phenomenon may be attributed to the students' perspectives on assessment tasks. Students tend to reject deeper approaches to study since the assessment involves a great deal of reproductive learning and they reckon that deeper approaches are not important or worth learning (Ramsden, 1998 & 1992; Entwistle, 1987 & 1992).

Perhaps, the students may not realize from their

hearts the benefit of the assessment task as learning task. When assessment tasks are correlated to learning tasks, the assessment should be at the hearts of those who: (a) learn, i.e., the student, (b) teach, i.e., the trainers, (c) hire, i.e., the employer, (d) develop the course or training program, i.e., the institution,

and (e) accredit the course or training program, i.e., the authority. All these people could be benefited from assessment. Brown et al. (1997) provides a list of the benefits of effective assessment practices and their beneficiaries (Exhibit 6):

Exhibit 6: Benefits of Effective Assessment Practice

Item	Benefit	To Student	To Teacher	To Employer	To Institution	To Authority
1	Providing feedback to the candidates to improve their learning.	✓	✓		✓	
2	Motivating the candidates.	✓	✓		✓	
3	Diagnosing a candidate's strengths and weaknesses.	✓	✓		✓	
4	Helping candidates to develop their skills of self-assessment.	✓	✓		✓	
5	Providing a profile of what a candidate has learnt.		✓		✓	
6	Passing or Failing a candidate.		✓		✓	
7	Grading or Ranking a candidate.		✓		✓	
8	Licensing the candidates to proceed.	✓				
9	Licensing the candidates to practice.	✓		✓		
10	Selecting candidates for future training programs.				✓	
11	Predicting the candidates' success in employment.			✓	✓	
12	Selecting the candidates for future employment.			✓		
13	Providing feedback to the trainers.		✓		✓	
14	Improving teaching.		✓		✓	
15	Evaluating the strengths and weaknesses of a training program.				✓	✓
16	Making a training program appear "respectable" and creditworthy to other institutions and employers.			✓	✓	✓

(Source: The subject research)

During the FGI, the respondents reached a consensus on the benefits of the assessment practices. Particularly, the examiners opined that 30 minutes' viva were inadequate for them to maximize the benefits. The registrars observed that some examiners habitually overran the viva by providing feedback to the candidates to improve their learning. If the examiners regarded the viva as a counseling session, motivating the candidates, diagnosing their strengths

and weaknesses, helping them to develop their skills of self-assessment, and even providing a profile of what a candidate has learnt, one could imagine how long a viva would take.

### 3.2 Assessment Task

Though the examiners' opinions have their grounds but the assessment task in the form of a viva could hardly afford it. Hence, the assessment



task of the subject program actually spread from the commencement of the action research to the attendance at the viva. However, nature of the assessment task evolved from the research process to the appraisal process. The task could be a Criterion-Referenced Assessment (CRA), Norm-Referenced Assessment (NRA) or Objective-Referenced Assessment (ORA).

Here comes a story from “Technical Literature”. Brown et al. (1997) quote the first ever “Assessment of SLO” was probably undertaken by Homo Australopithecus who once said to his son, “Now go out and kill your first bear.”. This hunting task appears to be an example of CRA in which the focus was placed on the outcome of “pass” or “fail”. If the hunting task is assigned to all the sons in the tribe and changed to: “Go out and kill as many bears as you can”, the focus of the assessment is shifted to the outcome of “bears hunted”. The number of bears hunted by each son would yield a rank order based on a distribution of scores. Then, the hunting task becomes a NRA.

The assessment task of the subject program can be termed a CRA rather than a NRA. The candidates are asked to identify a real-life problem in their workplaces and conduct an action research and compile a thesis which would be subject to a “pass-or-fail” rating system. When the candidates will not be ranked in the viva, they are not subject to a NRA.

However in the FGI, the Registrars reported that the thesis supervisors in the Mainland had a deep belief on NRA. They apparently believed that those candidates going to the viva should be placed on the top echelon. When the candidates were coming from the same institution but under the guidance of

different supervisors, the thesis supervisors tended to be leniently in rating the candidates under their own supervision. Understandably, the thesis supervisors guided the candidates to complete the research and the theses. They naturally rated their guided work as the most promising theses in the class.

From the perspectives of the internal and external examiners, they regarded the assessment task as an ORA instead of a CRA. During the FGI, the internal and external examiners reflected that the assessment task could tell whether the objectives of the subject program had been met. On this belief, they rated the candidates regardless to the overall distribution of scores. However, review of “Non-Technical Literature” surfaced that different panels of internal and external examiners had significant difference in rating a group of the candidates from the same class and same institution. This “CRA at Variance with NRA” phenomenon will be discussed later in this Paper.

Apart from the difference in rating amongst the thesis supervisors, the internal examiners and the external examiners, “Non-Technical Literature” also suggested a phenomenon under ORA. Regarding the assessment task as an ORA, the candidates and the examiners of the subject program appeared to have conflicting views on the objectives of the subject program. The candidates’ views on SLE versus the objectives of the subject program conflicting with the examiners’ views on SLO versus the objectives of the subject program. This phenomenon coincided with the notions of “fitness for purpose” and “fitness of award” that were reviewed in “Technical Literature”.

The links between these 2 notions and their inherent conflict in “educational quality assurance”

were covered in the data analysis pertaining to SLO. In this Paper, their implications for Assessment Practices are highlighted. The former notion examined the links between particular SLE and specific objectives of the subject program. In contrast, the latter notion concerned the links between the SLO against the national certification standards of Career Manager<sup>3</sup> through subject program.

Analysis of SLE revealed that the candidates had good learning experiences and believed that the objectives of the subject program were met upon their graduation. On the other hand, analysis of SLO disclosed that the examiners had reservation about the ASK of the candidates upon their graduation. Though the candidates successfully earned an MBA degree and gained the occupational title of Career Manager, there were still room for improvement in their ASK of being a Career Manager.

This “good SLE versus maybe better SLO” phenomenon in ORA replicated the subsistence of the notions of Aristotelian and Platonic interpretation of “good”. Aristotelian notion established that the subject program was “good for the candidates and for their career development”. “Good” was construed on the SLE of the candidates and their interpretation of “Fitness for Purpose” along with their perceived objectives of the subject program.

On the other hand, the internal and external examiners believed that “fitness of award” was independent of the candidates. Platonic notion established “maybe better SLO” should be construed on the basis of the pre-determined standards of the society. The examiners believed that there were ideal standards to which Career Managers should aspire. To this end, this research established that CRA may

be jeopardized by NRA whereas ORA may contribute to effective assessment practices in management education. Furthermore, candidates’ perception of “fitness for purpose” may contradict with the examiners’ perception of “fitness of award” under the Aristotelian and Platonic notions.

### 3.2.1 Variances of Assessment

In the preceding paragraph, a “CRA at Variance with NRA” phenomenon emerged. To understand more about this phenomenon concerning “reliability”, the Researcher randomly drew 5 from 23 panels. These 5 panels examined 72 candidates during the sampling period: 15 candidates by Panel 1, 14 candidates by Panel 2, 15 candidates by Panel 3, 14 candidates by Panel 4, and 14 candidates by Panel 5 (Exhibit 7).

Exhibit 7: Marking Variances amongst Supervisors, Internal & External Examiners of 5 Panels

Panel	Thesis Supervisor	Internal Examiner	External Examiner	Final Mark
1	75	67	69	70
	75	76	70	73
	73	66	69	68
	75	73	69	72
	75	75	71	80
	88	70	66	71
	75	67	66	68
	76	68	69	68
	78	64	68	66
	76	66	64	65
	75	73	68	74
	87	74	70	75
	75	73	70	71
	90	65	65	65
	92	70	73	74
2	72	72	65	68
	76	73	71	72
	74	66	70	67
	80	68	70	72
	76	65	68	0
	87	70	70	70
	80	74	79	68
	80	65	69	70
	75	71	68	74
	81	69	73	72
	83	60	66	68
	72	69	70	74
	68	68	64	68
	75	72	74	74

3. Chinese Career Manager Certificate (CCMC) is a professional national title granted by the National Certification Committee of the Chinese Career Manager, People's Republic of China. CCMC is a widely recognized vocational qualification for the recruitment, practicing, professional employment and development of Career Managers in China.

Panel	Thesis Supervisor	Internal Examiner	External Examiner	Final Mark
3	74	67	67	67
	81	64	69	66
	79	67	69	66
	76	64	57	67
	77	66	69	70
	82	70	70	72
	91	65	77	71
	78	65	67	62
	79	72	70	69
	75	63	68	65
	78	63	67	68
	73	72	69	71
	80	73	69	67
	83	68	74	71
	85	73	72	60
4	80	64	64	67
	78	66	70	70
	79	64	72	72
	75	65	70	67
	75	67	72	63
	78	65	72	62
	84	72	70	75
	73	68	69	76
	79	68	70	73
	85	64	68	64
	81	63	67	66
	77	62	59	55
	73	70	62	74
	87	65	69	76
	77	72	77	72
5	80	73	76	73
	88	64	71	70
	79	65	71	65
	78	68	70	72
	81	72	70	73
	80	75	79	74
	85	64	67	60
	78	64	70	70
	79	67	70	50
	80	75	68	50
	82	64	59	50
	80	68	69	70
	82	75	76	70

(Source: The subject research)

Amongst the 72 candidates, there are 5 failing cases: 1 in Panel 2, 1 in Panel 4, and 3 in Panel 5. They are excluded from the statistical analysis for homogeneity of variances. However, those failing cases (Exhibit 8) are studied successively in the following paragraphs.

Exhibit 8: Five Failing Cases of Viva Voce in Panel 2, 4 and 5

Failing Case	Thesis Supervisor	Internal Examiner	External Examiner	Final Score
1 (Panel 2)	76	65	68	0
2 (Panel 4)	77	62	59	55
3 (Panel 5)	79	67	70	50
4 (Panel 5)	80	75	68	50
5 (Panel 5)	82	64	59	50

(Source: The subject research)

In the first failing case, the panel discovered that the candidate only worked in the enterprise under study for one year as a supervisor. She was unable to answer the queries of the internal and external examiners about the research on “brand management”. Eventually, she admitted that one of her friends wrote the thesis for her. In the second failing case, the candidate was the director of public service organization. During the viva, she failed to prove that she had grasped the basic knowledge about “total learning organization” under study.

In the third failing case, the candidate conducted a questionnaire survey that had a strong theoretical foundation such as schools of thought of Marslow (Hierarchy of Need Theory) and Herzberg (Two Factor Theory). However, he could not answer the examiners’ queries about these theories in a satisfactory manner. In the fourth failing case, the candidate unreasonably used some outdated data from 2002-04 for SWOT (Strength, Weakness, Opportunity and Threat) analysis. Furthermore, she could not reason the use of BCG (Boston Consulting Group) Matrix in her study.

In the fifth failing case, the candidate had six years’ experience in property market. She submitted a thesis looked like a feasibility study on a new product in property market. The panel ruled that the theoretical base of the research and the application of knowledge and skills gained from the subject program to the real life problem appeared to be inadequate. In

five failing cases, the external examiners detected two of those sub-standard theses before the viva while the panels collectively detected another three during the viva.

The external examiners and academic panels appear to be essential instruments of fairness, reliability and validity. The presence of the external examiners and formation of panels were to protect the candidates and to safeguard standards. Protecting students implies checking on the fairness, reliability and validity of the assessment practice of the subject program. Safeguarding standards involves checking on the design of the assessment task, monitoring the SLO and making an estimate of worth of the subject program.

With the participation of external examiners, the collective assessment of the panel has particular advantages in upholding the reliability of the assessment practice of the subject program. McCormick (1979) relates reliability to the degree of relationship between or among the assessments of two or more independent assessors. For a sample of jobs, reliability is often measured by correlating pairs of independent assessments.

### 3.2.2 Findings of Statistical Tests

Previous studies (Scott, 1963) ascertained that the combination of the assessments of several people tended to increase the reliability of the composite assessments as long as all of them were good assessors. Reliability in this context is referred to the degree of relationship between or among the assessments of two or more independent assessors, or between separate assessments at different times by the same assessor. Reliability is measured by correlating pairs of independent assessment for a sample job, or

by correlating separate (test-retest) assessments made by the same assessor at different times. Assuming that the average test-retest reliability of an assessor is 0.80, the reliability coefficients for various numbers of assessors (Exhibit 9) could be:

Exhibit 9: Reliability Coefficients for Various Numbers of Assessors

Sample Size	1	2	4	6	8	16	20
Reliability	.80	.89	.94	.96	.97	.98	.99

(Source: Scott, 1963)

The pooled reliability of assessments tends to increase appreciably with even three or four assessors, and then increases more gradually (McCormick, 1979). Incidentally, there are hints that the pooled reliability of assessments made independently by the several assessors tends to be a bit higher than the reliability of assessments made collectively by panels of three to five assessors (Hoggatt and Hazel, 1970). These hints suggest that it is preferable to obtain individual assessor from two or more assessors and average them, rather than obtaining group assessments by consensus.

Panel may be advantageous in one way – independent assessment other than collective assessment. It is observed that the thesis supervisors, internal and external examiners are allowed to make individual assessment before the viva. Then, a panel comprising internal and external examiners makes the collective assessment after the viva. Such arrangement can be regarded as a good assessment practice of panel system.

However, it is noteworthy that there are variances between the markings of thesis supervisor, internal and external examiners on the 67 sampled candidates. If the final marks awarded by the academic panels

can be regarded as a “Control Group”, what are the variances between 3 individual assessments and the joint assessments made by the internal and external examiners in the viva?

Using SPSS 15.0 for Windows, the Contrast Coefficients of Assessments are worked out as per Exhibit 10: (1) Marks given by Thesis Supervisors, (2) Marks given by Internal Examiners, and (3) Marks given by External Examiners while (4) Final Marks awarded in the Viva by the Panels are regarded as a “Control Group”.

Exhibit 10: Contrast Coefficients of Assessments

Contrast	Assessment			
	1	2	3	4
1	1	0	0	-1
2	0	1	0	-1
3	0	0	1	-1

(Source: The subject research)

One way Analysis of Variance (ANOVA) test results are:  $F=94.918$ ,  $P=0.000<0.05$ . Statistically, there are significant variances in the assessments (Exhibit 11).

Exhibit 11: One Way ANOVA Test Results of the Marks given in 4 Assessments

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4952.966	3	1650.989	94.918	.000
Within Groups	4591.970	264	17.394		
Total	9544.937	267			

(Source: The subject research)

Test of Homogeneity of Variances (Exhibit 12) is conducted. The Levene statistics show that  $P=0.023<0.05$  does not assume equal variances. Therefore, further reference should be made to the “Does not assume equal variances” section of the Dunnett’s tD Test.

Exhibit 12: Test of Homogeneity of Variances in Assessments

Levene Statistic	df1	df2	Sig.
3.227	3	264	.023

(Source: The subject research)

Findings of Dunnett’s tD Test (Exhibit 13) indicate that the t value of Contrast 1 is 12.028 whereas  $P=0.000<0.05$ . Statistically, there are significant variances between the assessments made by (1) Thesis Supervisors and (4) Academic Panels. Contrast 2 with t value of -1.892 and  $P=0.061>0.05$  indicates that there are insignificant statistical variances between the assessments made by (2) Internal Examiners and (4) Academic Panels. Contrast 3 with t-value of -0.045 and  $P=0.064>0.05$  also indicates that there are insignificant statistical variances between the assessments made by (3) External Examiners and (4) Academic Panels.

Exhibit 13: Contrast Tests of Assessments by Thesis Supervisors, Internal and External Examiners

Contrast			Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Marks	Assume equal variances	1	9.42	.721	13.070	264	.000
		2	-1.28	.721	-1.781	264	.076
		3	-.03	.721	-.041	264	.967
	Does not assume equal variances	1	9.42	.783	12.028	126.014	.000
		2	-1.28	.678	-1.892	131.766	.061
		3	-.03	.667	-.045	131.199	.964

(Source: The subject research)

The findings of above priori comparisons suggest that there are significant variances between the assessments made by the thesis supervisors and the assessments made by the academic panels comprising internal and external examiners. Statistically, the assessments made individually or jointly by internal examiners and external examiners do not show significant variances. It will be clearer to study the

Honestly Significant Difference (HSD) of these assessments by using Newman-Kuels Studentized Range Test (S-N-K).

Findings of multiple comparisons (Exhibit 14) show that the mean difference of assessments made by the thesis supervisors at the 0.05 level when compared with the assessments made individually or jointly by the internal and external examiners.

Exhibit 14: Multiple Comparisons of Assessments

(Source: The subject research)

Multiple Comparisons							
		Mean					
Tukey HSD	1	2	10.701*	.721	.000	8.84	12.56
		3	9.448*	.721	.000	7.58	11.31
		4	9.418*	.721	.000	7.55	11.28
	2	1	-10.701*	.721	.000	-12.56	-8.84
		3	-1.254	.721	.305	-3.12	.61
		4	-1.284	.721	.285	-3.15	.58
	3	1	-9.448*	.721	.000	-11.31	-7.58
		2	1.254	.721	.305	.61	3.12
		4	-.030	.721	1.000	-1.89	1.83
	4	1	-9.418*	.721	.000	-11.28	-7.55
		2	1.284	.721	.285	-.58	3.15
		3	.030	.721	1.000	-1.83	1.89

Findings of S-N-K Test (Exhibit 15) show the difference between the assessments made individually or jointly by the internal and external examiners and prematurely by the thesis supervisors. All the samples (67 candidates) show significant difference. Tukey HSD Test (Exhibit 15) comes up with identical findings.

Exhibit 15: Post Hoc Multiple Comparisons of Assessments

Assessment	N	Subset for alpha = .05	
		1	2
Student-Newman-Keuls <sup>a</sup>	2	67	68.31
	3	67	69.57
	4	67	69.60
	1	67	79.01
	Sig.		.178
Tukey HSD <sup>a</sup>	2	67	68.31
	3	67	69.57
	4	67	69.60
	1	67	79.01
	Sig.		.285

Means for groups in homogeneous subsets are displayed.

<sup>a</sup> Uses Harmonic Mean Sample Size = 67.000.

(Source: The subject research)

Brown et al., (1997) suggest that the two main measures of reliability in assessment are measure of marking differences between examiners and within examiners. Historically, there has been plenty of evidence on the marking differences between examiners, even when using marking schemes. Exhibit 16 shows findings of previous research on marking differences:

Exhibit 16: Research and Findings on Marking Differences between Examiners from 1890 to 1994

Researcher	Previous Research	Findings
Edgeworth (1890)	Twenty-eight qualified examiners were invited to mark a Latin prose as if it were by candidates for the Indian Civil Service.	Marks ranged from 45 to 100 while the modal mark was 75.
Hartog and Rhodes (1935 & 1936)	Marking of English, history and chemistry papers in a school certificate examination.	Different examiners marked the same candidates as failed, passed or passed with credits.
Diederich (1957), Bell (1980), Newstead and Dennis (1994)	Fifty-three experts were invited to mark 300 short essays of year 1 university students.	All essays received five or more grades of the nine possible while 34 percent of the essays obtained all the grades.

(Source: The subject research)

In the subject program, assessment by different examiners produces considerable marking differences. Such differences influenced the fairness, reliability and validity of the assessment practice. In determining the appropriate marks, the part played by the supervisors or the examiners can be greater



than that of the student performance. However as compared with the previous research on marking differences, the variances in marking amongst the thesis supervisors, internal and external examiners appears to be nominal.

#### 4. Conclusion and Recommendations

With the analysis of other qualitative data, this research concludes that the candidates appear to have favorable SLE during the research process when there are proper guidance on the choice of research methodology and methods. Besides, they are happy that there is close supervision of the research activities. Simultaneously, the candidates perceive that guidance on the choice of research tools, compilation of thesis and literature are both relative unimportant and relevant unsatisfactory.

On the contrary, the examiners during the appraisal process perceive that these three aspects are relevant important. The candidates find that both the internal and external examiners are solemn in the appraisal process. They appreciate the professional knowledge of the internal examiners. They also feel that the viva voce enables them to proliferate their personal knowledge. All these give the candidates favorable SLE.

However, the candidates appear to have less favorable SLE in the appropriateness of the appraisal mode. It may be attributed to be different expectation between the candidate and the examiners on the SLO. When the examiners attack the candidates on their choice of research tools, literature review and compilation of the thesis, the candidates may have hard feeling toward the appropriateness of the appraisal mode. Such hard feeling may lead to the

candidates' relative unsatisfactory perception on the appropriateness, arrangement and environment of the viva voce.

Although the candidates originally perceive the above three aspects of viva voce as relative unimportant, their less favorable SLE may be intensified when the length of viva voce does not allow the candidates to have adequate interaction with the examiners during the viva. Some examiners tend to adopt "Tell and Sell" or "Tell and Listen" approach in the viva instead of "Problem-Solving" or "Composite Approach". This phenomenon reasons out that the candidates have favorable SLE in learning and research processes but less favorable SLE in the appraisal process.

Qualitative data analysis also reveals that the candidates of the subject program have a clear aim of being career managers. It is noted that some candidates work in public sector or stated-owned enterprises as high-ranking comrades or mid-career civil servants while some work in private sector or small-to-medium enterprises. They wish to pursue their career in the management profession. Completing the action research and passing the viva enables the candidates to get the national title of Chinese Career Manager Certification (CCMC).

In conducting the action research, some candidates may choose big topics because they are working in public sector or state-owned enterprises. Researches with large scope of study do not meet the requirements of the action research. Therefore, they have to revise their research projects in order to prove that they are fit for the award of an MBA degree and CCMC. Despite the obstacle in the form of a viva, the candidates still find the action learning and action

research of the subject program rewarding.

The examiners obviously believe that “MBA Candidate should be Genius Pig rather than Copy Cat”. The learning and research processes of the subject program aim to change the Attitudes, Skills and Knowledge (ASK) of the candidates and equip themselves to be career managers. During the appraisal process, the examiners diagnose the candidates’ ASK and assess their performance in tackling issues or problems in their action researches. In marking the candidates’ theses and assessing the candidates’ reflections, the examiners discover that some candidates appear to manifest “Descriptive Writing” or “Descriptive Reflection” rather than “Dialogic Reflection” or “Critical Reflection”.

The examiners diagnose that the candidates may simply copy what they have learnt from the subject program or the thesis supervisors know. The examiners long to see that the candidates of the subject program may become “understanding seekers” rather than “knowledge seekers”. Mere copying could hardly develop genuine “understanding seekers”. Qualitative data analysis suggests that understanding seeking candidates may have more desirable SLO than knowledge seeking candidates.

Candidates with desirable SLO will be awarded an overall grading of “outstanding” in their assessment. However, only a great minority of the candidates can earn the highest overall grading in their action researches, master theses and viva performance. This type of observable SLO, namely “Extended Abstract” is scarce in the subject program. Instead, lower levels of observable SLO like “Relational, Multistructural, Unstructural or Pre-Structural” are predominant amongst the candidates of the subject program.

It is preferable for a candidate to be a genius pig than a copy cat. The worst copy cats are those candidates stealing other people’s intellectual property or buying theses for submission as their own works. Some obvious misconduct in referencing protocol may be attributed to the unfamiliarity with the academic conventions. Through the appraisal process of the subject program, examiners could assess which candidates are copy cats and which candidates are genius pigs.

Besides, the thinking, learning and writing styles of the candidates may be affected by their background, mindsets and even their organizational climate. Some candidates are nurtured in planned economy or politician job settings which influence their learning orientations, mindsets and writing styles. They are used to a paradigm of political writing which annoys the examiners. This phenomenon is once popular, particularly amongst those candidates working in public sector or state-owned enterprises.

Background, mindsets or organizational climate may also affect the SLO in the aspects of the candidates’ ASK. Qualitative data analysis also suggest a correlation between the candidates’ ASK and “Understanding” as reported in the preceding paragraphs. The correlation of these four learning components is profound. Nonetheless, this research surfaces that the desirable SLO should be a common output of these four learning components.

Output of the above four learning components could develop the candidates’ cognitive, social and cognitive skills. These three skills are essential to the personal growth and career development of the candidates. In the subject program, the skills of the

candidates are taught through the medium of the core and elective subjects and the guidance of the thesis supervisors. Ironically, the thesis supervisors assess that the candidates have grasped these three skills before the viva. The examiners may think the other way in the viva.

As a result, “Supervisor’s Assessment may contrast with Examiner’s Assessment”. Qualitative data analysis reveals that the thesis supervisors had a deep belief on NRA. Therefore, they tend to overmark the candidates before the viva. In contrast, the examiners regard the assessment task as an ORA instead of a CRA. They tend to assess the candidates’ performance against the objectives of the subject program. Therefore, there are variations between the supervisors’ assessment and the examiners’ assessment. Quantitative data analysis also sheds light to this phenomenon.

In retrospection, the history of significant variations between the supervisors’ assessment and examiners’ assessment is traced. Such phenomenon has subsisted in the assessment practices of the subject program since 2003. Statistical tests elaborate this phenomenon in an explicit way. Quantitative data analysis suggests that there are considerable variances between the markings of thesis supervisor, internal and external examiners on the sampled candidates.

Further study of technical literature discloses that there is a long history of marking variances between assessors from 1890. In comparison, the variances in marking amongst the thesis supervisors, internal and external examiners of the subject program appear to be nominal. Nonetheless, this research substantiates that “variances in marking” is one of the areas to be

improved in the assessment practices of the subject program.

Qualitative data analysis of this research reveals that the overall SLE of the candidates on the subject program is favorable. As read with the findings of the quantitative data analysis, it is evident that the less favorable SLE stems from the candidates’ perception on the “fairness” and “reliability” of the assessment task. Marking variances between the thesis supervisors and the examiners are contributory to this phenomenon.

Qualitative data analysis also discloses that some candidates with favorable SLE do not achieve desirable SLO. From the perspectives of the examiners, the observable SLO of most candidates are far from the ideal state such as “Understanding Seekers” or “Prospective Learners”. Therefore in the subject program, the candidates’ favorable SLE cannot equate with the desirable SLO.

This research further surfaces a “good SLE versus maybe better SLO” phenomenon. The candidates consider the subject program a good one on their SLE and their interpretation of “Fitness for Purpose” along with their perceived objectives of the subject program. On the other, the examiners consider the subject program on the basis of the pre-determined standards of the society and their interpretation of “Fitness of Award” along with the performance of the candidates in the assessment tasks. Conflicting views of the candidates and the examiners are also contributory to the emergence of the grounded theory.

The resultant Grounded Theory is “Robust Assessment Practices can ensure Quality Education”. This research reveals that robust assessment practices

are built on fairness, reliability and validity. On the other, robust assessment practices can be built on effective assessment tasks taking a sample of what the candidates do, making references and estimating the worth of their actions. This research surfaces some shortcomings of the above assessment task undertaken by the candidates of the subject program:

- ✧ Overload of the candidates and the examiners.
- ✧ Insufficient time for the candidates to complete the action research in the time available and insufficient time for the examiners to mark the theses before the viva.
- ✧ Inadequate or superficial feedback provided to the candidates during the viva.
- ✧ Wide variations in assessment demands of different panels and marking across panels or within a panel.
- ✧ Wide variations in marking by supervisors.
- ✧ Fuzzy or non-existent criteria; undue precision and specificity of marking schemes or criteria.
- ✧ Candidates do not know what is expected of them or what counts as a good or bad research or thesis.

Despite the above shortcomings, the examiners and registrars of AIOU have a strong belief on effective assessment practices which could benefit all the stakeholders of the subject program. To ensure the effectiveness of assessment practices in the subject program, there is a robust assessment system involving thesis supervisors, internal and external examiners in place. Although there are variations in the inferences drawn by the different examines and different panels, AIOU has placed extra efforts in ensuring quality of the subject program through assessment practices.

In estimating the worth of the candidates' action in their research projects, the internal and external examiners play an important quality role to ensure

the “fairness”, “reliability” and “validity” of the assessment task. Apparently, the four-tier markings of the candidates by thesis supervisors, internal examiners, external examiners and academic panels are robust assessment practices for quality assurance of the subject program. The appraisal process in the form of a viva also plays a critical role of gatekeeper for quality assurance.

## Reference

- Atkins, M. J., Beattie, J., & Dockrell, B. (1993). *Assessment Issues in Higher Education*. Sheffield: Employment Department.
- Bain, G., & Thomas, P. (1984). Contextual Differences of Learning Approaches: The Effects of Assessment. *Human Learning* (3), 227-240.
- Barnett, R. (1997). *Realising the University*. London: University of London, Institute of Education.
- Bell, R. C. (1980). Problems in improving the Reliability of Essay Marks. *Assessment and Evaluation in Higher Education*, 5, 254-263.
- Biggs, J. (1987). *Student Approaches in Learning and Studying*. Melbourne: Australian Council for Educational Research.
- Blake, R. L., & Vernon, D. T. A. (1994). Does Problem-Based Learning Work? A Meta-Analysis of Evaluative Research. *Academic Medicine*, 69, 550-563.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing Student Learning in Higher Education*. London & New York: Routledge.
- Clarke, R., & Newble, D. I. (1987). *Approaches to Learning in a Traditional and an Innovative Medical School*. Milton Keynes: SRHE & Open University.
- Diederichm, P. (1957). *The Improvement of Essay Examinations*. Princeton: Educational Testing Service.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of Educational Measurement*. New York: Prentice Hall.
- Edgeworth, F. Y. (1890). The Elements of chance in Competitive Examinations. *Journal of the Royal Statistical Society*, 400-475, 644-463.
- Entwistle, N. J. (1987). *Styles of Learning and Teaching* (2nd ed.). Chichester: Wiley.
- . (1992). *The Impact of Teaching on Learning Outcomes* (Rev. ed.). Sheffield: CVCP UCoSDA.
- Eysenck, M. W., Piper, D. W., & Richardson, J. T. E. (1987). *Student Learning, Research in Education and Cognitive Psychology*. Milton Keynes: SRHE & Open University.
- Gronlund, N. E. (1988). *How to construct Achievement Tests*. New York: Prentice Hall.
- Haper, C., & Kember, D. (1989). Interpretation of Factors

- Analyses from the Approaches to Studying Inventory. *British Journal of Educational Psychology* (59), 66-74.
- Hartog, P., & Rhodes, E. C. (1935). *An Examination of Examinations*. London: Macmillan.
- . (1936). *The Marks of Examiners*. London: Macmillan.
- Hattie, J., & Watkins, D. (1983). A Longitudinal Study of the Approaches to learning of Australian Tertiary Students. *Human Learning* (4), 127-141.
- Hoggatt, R. S., & Hazel, J. T. (1970). *Reliability of Individual versus Group Job Pay Rating*. Texas: USAF.
- McCormick, E. J. (1979). *Job Analysis: Methods and Applications* (1st ed.). New York: AMACOM.
- Newstead, S., & Dennis, I. (1994). Examiners examined: The Reliability of Exam Marking in Psychology. *The Psychologist*, 7, 216-219.
- Ransden, P. (Ed.). (1988). *Improving Student Learning*. London: Kogan Page.
- Rowntree, D. (1987). *Assessing Students - How shall we know them?* London: Harper & Row.
- Scott, W. E. J. (1963). *The Reliability and Validity of a Six-Factor Job Evaluation System*. New York: Purdue University.

## Authors' Profile

Gryphon Sou earned a Doctor of Management Degree from the International Management Center – Southern Cross University, Australia and a Post-Doctoral A+ Enhancement Award from IMC – Revans University, UK/US. He is now working as a Visiting Professor in the Asia International Open University (Macau) and is pursuing a Doctor of Education Degree from University of Technology Sydney. He is a Life Member of Hong Kong Teachers Association as well as a Fellow of The College of Preceptors and The College of Teachers, UK.  
e-mail: [GryphonSou@AIOU.edu](mailto:GryphonSou@AIOU.edu)

仇志成持有美國加州海岸大學工程理學士、澳洲天主教大學行政學碩士及國際管理中心—澳洲南十字星大學管理學博士學位，他在英國伯明翰郡國際管理中心—美國科羅拉多州銳文大學完成博士後研究，獲取優等文憑。他現任亞洲（澳門）國際公開大學客座教授，同時在澳洲雪梨科技大學攻讀教育學博士學位。他亦是香港教師會終生會員、英國師範學院院士及英國教師學會深資會員。

Received: 13.8.08, accepted 8.9.08, revised 19.10.08